ShuffleFaceNet：高效轻巧的人脸识别轻巧的人脸架构

Yoanna Mart′ınez-D′ıaz，Heydi Mendez- V′azquez，Miguel Nicol′ as- D′′ıaz

先进技术应用中心（CENATAV）

7A *♯* 21406 西波涅，普拉亚，PC 12200，哈瓦那，古巴

{ ymartinez，hmendez，mnicolas } @ cenatav.co.cu

路易斯·卢瓦诺（Luis S.Luevano），李奥纳多·张（Leonardo Chang），米格尔·冈萨雷斯（Miguel Gonzalez-Mendoza）

蒙特雷理工学院（Tecnologico de Monterrey），工程与科学学院，

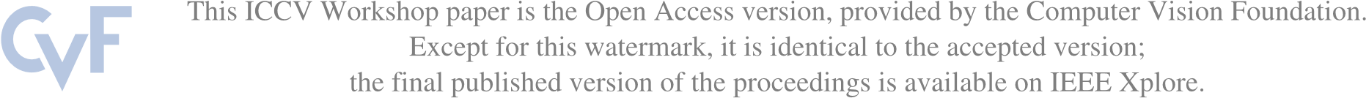
墨西哥

luisluevano@outlook.com， { lchang，mgonza } @ tec.mx

# 摘要

*卷积神经网络的最新成功导致了各种新的有效架构的发展。但是，针对面部识别的特定情况而设计的却很少。受到最新的ShuffleNetV2模型的启发，本文提出了一种轻量级的面部结构。该提案名为ShuffleFaceNet，对其进行了重大修改，以提高人脸识别的准确性。在相同的实验条件下，ShuffleFaceNet的准确性要比原始ShuffleNetV2显着提高，并保持相同的速度和紧凑的存储空间。此外，*

# 简介

~~深度神经网络（DNN）最近在许多计算机视觉任务中取得了一系列突破，包括无约束的人脸识别[ 33 ]。然而，现代高度精确的面部识别方法通常在非常深的卷积神经网络（细胞神经网络）[内置6 ，26 ，39 ]，这意味着它们包括卷积层的长序列。结果，这些模型需要大量的计算资源，例如大内存和强大的GPU，才能获得高性能的结果。例如，广泛使用的VGGFace网络[ 26 ]涉及1.38亿个参数，超过500 MB的内存大小以及超过15G的浮点运算（FLOP）来对单个224进行分类× 224图像。因此，这种模型通常无法部署在实时应用程序或计算受限的平台上，例如机器人，智能手机和移动设备~~。

为了克服这些限制，最近的发展集中在构建小型高效的神经网络而又不显着降低其性能的情况下。~~一些方法已经尝试通过使用诸如修剪[ 9 ]，知识蒸馏[ 11 ]，低秩逼近[ 41 ]和量化[ 16 ]之类的技术来压缩或加速预训练网络。在过去的几年中，开发轻便深层神经网络是最有前途的解决方案，以获得一个更好的速度精度权衡[ 14 ，40 ，19 ，27 ]。SqueezeNet [ 14 ]，MobileNets [ 27 ]和ShuffleNets [40 ，19 ]是最流行的常见视觉识别任务中，表现出了不俗的业绩。然而，短短几年的作品提出了专门针对人脸识别[设计轻巧准确架构2 ，4 ，7 ，36 ，37 ]，因此这个话题值得进一步关注。~~

在本文中，我们提出了一种名为ShuffleFaceNet的新型轻量级体系结构，该体系结构将效率极高的网络ShuffleNetV2 [ 19]扩展到了人脸识别领域。首先，我们将全局平均池化层替换为全局深度卷积层，以获得更具区分性的人脸表示。其次，由于其得到的精度比整流线性单位（ReLU）函数更高，因此我们将参数整流线性单位（PReLU）用作非线性激活函数。结果，我们设计了四个具有不同复杂度级别的ShuffleFaceNet模型，这些模型使用少于450万个参数，最大计算复杂度和模型大小分别为1.05G FLOP和18 MB。在图像和视频基准数据集上进行的实验表明，我们的建议同时改进了最先进的轻量级和非常深的CNN，可用于人脸识别。这项工作的主要贡献概述如下：

* 通过使用全局深度卷积层输出判别特征向量和PReLU作为非线性激活函数，我们不仅确保了计算速度和紧凑的存储空间，而且还显着提高了人脸识别精度。
* 我们设计了一种高效，准确的轻型人脸架构，具有四个不同的复杂性级别。生成的ShuffleFaceNet模型的大小小于20 MB，实际推理CPU时间约为37 ms，适用于在实时应用程序以及移动和嵌入式设备上进行部署。
* 我们证明，在两个流行的人脸识别基准以及最近发布的轻量级人脸识别挑战中，该提案比最先进的轻量级CNN和非常深的CNN都具有更好的性能。

本文的结构如下。第2节回顾了用于面部识别的现有轻量级CNN。第三部分介绍了建议用于面部识别的轻量级ShuffleFaceNet架构。实验结果在第4节中给出，最后，我们在第5节中得出结论。

# 相关工作

~~部署有效，轻量级的深度面部识别架构仍然是现实应用程序的挑战。在[ 37]以获得紧凑且低维的面部表示。为此，作者介绍了ReLU函数的替代方法，称为Max-Feature-Map（MFM），以抑制每个卷积层中的低激活神经元。另外，小型卷积滤波器，网络中的网络层和残差块用于减少参数空间并提高性能。评估了轻型CNN的三种体系结构，与最新的大脸模型相比，在速度和存储空间方面显示出更好的性能。但是，它们最精确的架构（轻型CNN-29）具有1260万个参数和大约3.9G FLOP，这使其不适用于移动和嵌入式平台。[ 36中提出了一种空间卷积的替代方法，称为移位运算。]，它需要零个FLOP和零个参数，并且可以轻松有效地实现。为了证明移位操作在面部识别任务中的有效性，作者采用了原始的FaceNet [ 28 ]，并提出了一个新的ShiftFaceNet模型，该模型将参数大小减小了35 × ，在上述三个验证基准中，精度下降了2％。LMobileNetE [ 4 ]是MobileNet [ 27 ] 的改进版本，以最低的运行时间实现了可比的人脸验证准确性，但是模型大小为112MB，这是大型模型，而不是轻量级模型。~~

~~最近，MobileFaceNets [ 2 ]被引入，用于在移动和嵌入式设备上进行高精度和实时人脸验证。在Wild（LFW）数据库中的“带标签的面孔”上进行的实验表明，MobileFaceNets可以达到与最新的大型模型相似的准确性，并且推理速度更快。但是，对于MegaFace数据集，准确性会略有下降。MobiFace [ 7 ]是为移动设备上的人脸识别而设计的另一种轻量级CNN，它采用快速下采样和具有扩展层的瓶颈残留块，并在LFW数据库上达到99.7％，在MegaFace数据库上达到91.3％的高性能。~~

~~无论作品很少集中在专为脸部识别设计的高效轻巧的架构，还有其他轻量级细胞神经网络[ 14 ，40 ，19 ]这表明，在图像分类任务性能优良，值得进一步关注人脸识别。例如，SqueezeNet [ 14 ]是一个非常小的CNN架构，它使用少50 倍的参数就可以在ImageNet上实现AlexNet级的准确性。它的成功取决于三个主要策略：首先，用1 × 1过滤器替换3 × 3过滤器，该过滤器的参数减少9 × 。第二，将输入通道数减少到3 ×3个过滤器，最后在网络中进行下采样，以使卷积层具有较大的激活图。张等。[ 40 ]提出了一种名为ShuffleNet的计算效率极高的CNN架构，该架构利用了逐点分组卷积和信道随机播放操作。与MobileNet相比，它在ImageNet分类上获得了可观的利润。ShuffleNetV2 [ 19 ]受到ShuffleNet的启发，但考虑到其设计中的一些实际方面，以在保持高水平准确性的同时获得更有效的网络体系结构。为此，引入了一个简单的运算符，称为通道拆分，它可以保持大量且同等宽的通道，而无需密集卷积或过多的组。~~

# ShuffleFaceNet架构

在本节中，我们将详细介绍为面部识别而设计的轻量级ShuffleFaceNet架构。该提案的灵感来自于最新的网络ShuffleNetV2 [ 19 ]，但增加了一些旨在提高其在人脸识别任务上的鲁棒性的策略。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 名称 | 内核/步幅 | 输出尺寸 | 输出通道 | | | | | 0 *。*5× | 1× | 1 *。*5× | 2× | | 图片 | -- | 112 × 112 | 3 | 3 | 3 | 3 | | 转换1 | 3×3 */* 2 | 56 × 56 | 24 | 24 | 24 | 24 | | 第二阶段 | -- | 28 × 28 | 48 | 116 | 176 | 244 | | 第三阶段 | -- | 14 × 14 | 96 | 232 | 352 | 488 | | Stage4 | -- | 7 × 7 | 192 | 464 | 704 | 976 | | 转换5 | 1×1 */* 1 | 7 × 7 | 1024 | 1024 | 1024 | 2048 | | GDConv | 7 × 7/1 | 1 × 1 | 1024 | 1024 | 1024 | 1024 | | 线性转换 | 1 × 1/1 | 1 × 1 | 128 | 128 | 128 | 128 |   表1. ShuffleFaceNet架构具有四个不同级别的复杂性。 |

大多数用于图像分类的深度网络（包括ShuffleNetV2）在嵌入过程中都使用全局平均池（GAP）层的输出作为特征向量。然而，在人脸识别的情况下，这种策略已证明是较不准确的[ 37 ， 2 ，4]。这是由于GAP层平等地处理输出特征图的每个单元，这与不同的单元带来或多或少的区分性信息以提取面部特征向量的理论不一致。相反，使用全连接（FC）层使我们可以了解这些单元的不同权重。然而，FC层最终具有大量权重，这不仅需要更多的计算能力，而且会增加模型大小。最近，全局深度卷积（GDC）层用于[ 2]处理具有不同重要性的输出特征图的不同单元，显示出一种有效的人脸识别结构。在这项工作中，我们将ShuffleNetV2的GAP层替换为GDC层。

另一方面，ShuffleNetV2模型基于ReLU激活函数[ 24 ]，该函数通常提供高维和稀疏特征。为了缓解这一问题，一些激活已经被提出[ 37 ，10 ，20 ，38 ]。我们选择PReLU [ 10 ]作为非线性而不是RELU这已被证明是用于面部识别[更好1 ，2 ，7 ]，因为它允许否定响应，在又改善了网络性能。

另外，我们在网络的开头使用快速下采样策略，并在GDC层之后使用线性1×1 卷积层作为特征输出层。结果，获得了紧凑的128维面部表示。

提议的ShuffleFaceNet体系结构的详细结构如表1 所示，每个块中的通道数按比例缩放以生成不同复杂度的网络，表示为0 .5× ，1× ，1 .5× 和2× 。阶段2-4中的构建块由DenseNet块[ 12 ]组成。

# 实验评估

在本节中，我们将从准确性，速度和模型大小方面评估轻量级ShuffleFaceNet架构的性能。将该提案与三个用于人脸识别的基准数据集上的几个最新模型进行了比较。

## 训练和网络设置

我们使用MS1M数据集[ 8 ] 作为训练集，其中包含510万张93K个身份的人脸图像。我们将带有标签的野外面孔（LFW）[ 13 ]，正面人物（CFP）[ 29 ]和年龄数据库（AgeDB）[ 23 ]作为验证集。使用RetinaFace检测器[ 5 ]将所有面部图像重新对齐为112 × 112 的尺寸，并通过减去127.5然后除以128来归一化RGB图像中的每个像素（范围在[0; 255]之间）。所有特征嵌入尺寸设置为128。

我们将批量大小设置为256，并在两个Nvidia GeForce GTX 1080Ti（11GB）GPU上训练模型。学习率从0.1开始，并在100K，140K，160K迭代中除以10。总迭代步骤设置为200K。我们使用了随机梯度下降优化器，将动量设置为0.9，重量衰减设置为5e-4。卷积的参数初始化为Xavier，并从高斯正态分布中随机采样。所有实验均在MxNet框架上进行[ 3 ]。

我们用四个复杂度级别和不同的损失函数（例如SoftMax，CosFace [ 32 ]和ArcFace [ 4 ]）训练了ShuffleFaceNet 。表2 显示了在LFW，CFP-FP和AgeDB数据集上为每种模型获得的验证准确性，以及参数的数量，模型的大小以及FLOP的复杂性。

如我们所见，在LFW的情况下，具有不同复杂度级别的每个ShuffleFaceNet的性能改进并不显着，因为该数据集几乎已饱和。相反，CosFace和ArcFace的性能均优于SoftMax，尤其是在较大的姿势和年龄变化下。此外，我们发现CosFace和ArcFace的性能非常相似。但是，最后一个要好一些。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 方法 | 复杂 | ＃参数。 | 型号尺寸 | 失利 | 准确性 | 准确性 | 准确性 | |  | 襟翼 | （男） | （MB） | 功能 | LFW | 计划生育计划 | 年龄数据库 | |  |  |  |  | SoftMax | 96.87 ± 1.2 | 88.60 ± 1.4 | 78.33 ± 2.0 | | ShuffleFaceNet 0 *。*5× | 6690万 | 0.5 | 1.9 | CosFace | 99.23 ± 0.5 | 92.59 ± 1.4 | 93.22 ± 1.4 | |  |  |  |  | 弧面 | 99.07 ± 0.5 | 91.87 ± 1.5 | 92.45 ± 1.7 | |  |  |  |  | SoftMax | 96.91 ± 0.8 | 90.42 ± 2.0 | 80.40 ± 1.5 | | ShuffleFaceNet 1× | 275.8百万 | 1.4 | 5.6 | CosFace | 99.42 ± 0.3 | 95.07 ± 0.7 | 95.13 ± 1.0 | |  |  |  |  | 弧面 | 99.45 ± 0.4 | 96.04 ± 0.9 | 96.33 ± 0.7 | |  |  |  |  | SoftMax | 96.37 ± 0.9 | 90.57 ± 1.0 | 80.37 ± 2.7 | | ShuffleFaceNet 1 *。*5× | 577.5百万 | 2.6 | 10.5 | CosFace | 99.62 ± 0.2 | 96.79 ± 0.6 | 96.75 ± 0.7 | |  |  |  |  | 弧面 | 99.67 ± 0.3 | 97.26 ± 0.7 | 97.32 ± 0.8 | |  |  |  |  | SoftMax | 97.15 ± 0.7 | 91.57 ± 1.0 | 81.95 ± 2.9 | | ShuffleFaceNet 2× | 1.05克 | 4.5 | 18岁 | CosFace | 99.58 ± 0.3 | 97.33 ± 0.6 | 97.08 ± 0.9 | |  |  |  |  | 弧面 | 99.62 ± 0.4 | 97.56 ± 0.6 | 97.28 ± 0.8 |   表2. LFW，CFP-FP和AgeDB数据库上不同损失函数的验证结果（％）。   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | 方法 | LFW | 计划生育计划 | 年龄数据库 | 复杂 | #Params | 型号尺寸 | GPU速度 | |  | 准确性 | 准确性 | 准确性 | （FLOP） | （男） | （MB） | （女士） | | ShuffleNetV2 \* 1 *。*5× | 99.52 ± 0.4 | 96.21 ± 1.1 | 94.78 ± 1.1 | 577.3百万 | 2.5 | 10.1 | 0.77 | | ShuffleFaceNet 1 *。*5× | 99.67 ± 0.3 | 97.26 ± 0.7 | 97.32 ± 0.8 | 577.5百万 | 2.6 | 10.5 | 0.77 |   表3.在LFW，CFP-FP和AgeDB数据库上，我们的ShuffleFaceNet和原始ShuffleNetV2之间的性能比较。 |

如预期的那样，复杂度越高，精度越高。但是，级别1 之间没有显着差异*。*5× 和2× 。因此，为了获得更好的速度精度折衷，我们决定在剩余的实验中使用经过ArcFace损失函数训练的ShuffleFaceNet 1.5 × 。

### 与ShuffleNetV2架构的比较

我们将ShuffleFaceNet 1.5 × 与原始ShuffleNetV2 1.5 × [ 19 ]进行比较，以显示该方案在人脸识别方面的优势。为了公平起见，在与我们的ShuffleFaceNet相同的训练设置下，通过ArcFace损失功能从零开始训练ShuffleNetV2 1.5 × 。在本文的其余部分中，我们将所得模型称为ShuffleNetV2 \* 1.5 × 。

表3 列出了LFW，CFP-FP和AgeDB数据集上测试模型的验证准确性。此外，它还显示参数数量，模型大小和推断时间。可以看出，该提案在三个数据集上的表现均优于原始ShuffleNetV2模型。另一方面，尽管模型的大小和参数数量有所增加，但推理时间仍保持不变。这意味着维护了高效ShuffleNetV2设计中考虑的准则[ 19 ]。

## 绩效评估

为了评估ShuffleFaceNet 1.5 × 的有效性，我们在两个流行的基准以及最近发布的“轻量级面部识别挑战”上进行了几次实验。请注意，我们不会在测试数据库的任何训练集上对ShuffleFaceNet模型进行重新训练或微调。因此，我们直接提取在前述清洁MS1M数据集上学习到的ShuffleFaceNet的特征，并通过度量标准对这些特征进行比较。

### FaceScrub上的MegaFace挑战1

|  |  |  |
| --- | --- | --- |
| 方法 | 等级1 | VR @ FAR = 10 -6 |
| Vocord-DeepVo1 | 75.1 | 67.3 |
| 深度感大 | 74.8 | 87.8 |
| 中心损耗[ 34 ] | 65.2 | 76.5 |
| 脸网[ 28 ] | 70.5 | 86.5 |
| 脸部美容[ 32 ] | 82.7 | 96.7 |
| ResNet50-ArcFace [ 4 ] | 77.5 | 92.3 |
| ResNet100-ArcFace [ 4 ] | 81.0 | 97.0 |
| 轻型CNN-4 [ 37 ] | 60.2 | 62.3 |
| 轻型CNN-9 [ 37 ] | 67.1 | 77.5 |
| 轻型CNN-29 [ 37 ] | 73.5 | 84.7 |
| MobileFaceNet [ 2 ] | -- | 90.2 |
| ShuffleNetV2 \* 1 *。*5× | 69.6 | 84.1 |
| ShuffleFaceNet 1 *。*5× | 77.4 | 93.0 |

表4.使用FaceScrub作为测试集的MegaFace Challenge 1的性能评估。等级1是指具有第一百万个干扰物的第一位置的面部识别精度（％），VR（％）对应于FAR值为10 -6的验证TAR 。

MegaFace数据库[ 17 ]是最大的公开测试基准之一，用于评估百万分之一的人脸识别算法的性能。MegaFace包括一个画廊套件和一个探针套件。画廊集由Yahoo [ 31 ] 的Flickr照片子集组成，包含来自69万个不同个人的一百万多幅图像。探针集是两个现有数据库：FaceScrub和FGNet。在这项工作中，我们使用FaceScrub [ 25 ]作为探针集，其中包含530个独特个体的100K张照片。

表4 总结了由拟议的ShuffleFaceNet 1.5 × 和原始ShuffleNetV2 \* 1 获得的结果*。*与报告用于识别和验证任务的最新方法相比提高了5倍。在错误接受率（FAR）为10 -6下的正确接受率（TAR）用于报告验证结果，而在识别情况下将采用Rank-1面部精度。由于我们的训练集有超过50万张图像，因此被认为是很大的。

我们可以看到，在大型数据库上，ShuffleFaceNet 在等级1和VR @ FAR = 10 -6评估指标上均比ShuffleNetV2 \*几乎高出9％，并显示出其相对于其他轻型模型（如MobileFaceNet和Light）的优越性。 CNN -4，-9和-29。获得的结果甚至比现有的非常深的模型（例如Vocord和Deepsense）要好，后者作为基准的基准方法提供[ 17 ]。其他能够提供更好结果的非常深入的模型需要更高的计算资源和存储空间，我们将在后面进行分析。例如，在此基准数据库中获得最佳结果的LResNet100E模型的大小为250 MB。

### 在YouTube Face数据库上进行评估

YouTube面孔（YTF）数据库[ 35 ]是一个大型视频数据集，用于在视频中进行无限制的人脸识别。它包含1,425个主题的3,425个视频，这些视频在表情，照明，姿势，分辨率和背景上都有很大差异。每个主题平均提供2.15个视频。视频的平均长度为181.3帧。对于YTF数据库的标准协议，提供了对应于5000个视频对的配对匹配基准。具体而言，将这些对分为十对，每个对包含250个正对和250个负对。

对于每个YTF视频，我们选择了50个最正面的帧并计算相应的面部描述符。最后，视频由50个面部描述符的平均值表示，并且余弦相似度用于比较。

我们比较了SuffleFaceNet 1 的性能*。*使用ShuffleNetV2 \* 1的5倍*。*该数据库报告了5倍和最新技术。表5 以通常被认为是报告验证结果的三个指标来表示获得的验证结果：平均准确度，曲线下面积（AUC）和均等错误率（EER）。可以看出，该提案获得了最新的结果，并且在这种情况下，对于三个使用的指标，其性能均优于原始ShuffleNetV2 \*。

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | 准确性 | AUC | 能源效率 |
| LBinVF 2 [ 21 ] | 83.3 | 93.2 | 14.6 |
| DeepFace-单身[ 30 ] | 91.4 | 96.3 | 8.6 |
| ShiftFaceNet [ 36 ] | 90.1 | 96.1 | -- |
| 南[ 39 ] | 95.7 | 98.8 | -- |
| 脸网[ 28 ] | 95.1 | -- | -- |
| 轻型CNN-29 [ 37 ] | 95.5 | -- | -- |
| 球面[ 18 ] | 95.0 | -- | -- |
| VGG脸[ 26 ] | 97.3 | -- | -- |
| 脸部美容[ 32 ] | 97.6 | -- | -- |
| 中心损耗[ 34 ] | 94.9 | -- | -- |
| TBE-神经网络[ 6 ] | 95.0 | -- | -- |
| ShuffleNetV2 \* 1 *。*5× | 93.3 | 97.7 | 7.0 |
| ShuffleFaceNet 1 *。*5× | 95.7 | 98.2 | 5.3 |

表5. YouTube Face数据库上的验证结果（％）。

最近，针对YTF数据库提出了新的相关评估协议：REP-YTF [ 22 ]，其中包含视频/视频比较和视频/图像比较的开放/封闭集标识。在这些协议下，YTF数据库分为十个训练和测试集的随机试验。在每个试验中，对于开放式和封闭式识别协议，都通过使用开放度值获得测试集的三种不同配置：0.2、0.5和0.9，从而产生不同的画廊尺寸。在十个随机试验中计算性能指标并取平均值，并报告标准偏差。

我们比较ShuffleFaceNet 1 *。*5× 和ShuffleNetV2 \* 1 *。*通过使用开放式和封闭式识别协议，在REP-YTF上报告的三种表现最佳的方法达到5倍 。基于在[ 22 ] 上获得的结果，我们选择LDA度量学习来进行比较。对于开放式识别协议，表6 显示了在10项试验中，视频转播到视频转译的错误接受率（FAR）为1％的情况下，第1级的平均检测和识别率（DIR）以及相应的标准差。视频和视频到图像方案。在闭集识别协议的情况下，表7中列出了等级1的识别率 。如我们所见，在此协议下，ShuffleFaceNet和ShuffleNetV2 \*之间的差异通常也约为10％，并且两者都大大超过了基线方法。例如，在视频到图像的开放场景中，ShuffleFaceNet的准确性比流行的VGG-Face模型高出40％以上。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 方法 | 视频到视频 | |  |  | 视频到图像 | | | 运算（0.2） | 运算（0.5） | 运算（0.9） | 运算（0.2） | 运算（0.5） | 运算（0.9） | | LBinVF 2 + LDA | 10.05 ± 2.1 | 8.57 ± 0.8 | 8.18 ± 1.0 | 6.58 ± 1.5 | 4.78 ± 0.8 | 4.53 ± 0.5 | | VGG-Face + JB | 22.83 ± 3.6 | 18.16 ± 1.8 | 16.28 ± 1.5 | 17.33 ± 2.9 | 14.20 ± 2.4 | 13.14 ± 1.1 | | Dlib + LDA | 25.97 ± 3.0 | 20.12 ± 1.2 | 17.99 ± 1.5 | 16.62 ± 4.2 | 14.26 ± 1.7 | 11.41 ± 1.0 | | ShuffleNetV2 \* 1 *。*5× + LDA | 51.63 ± 4.7 | 45.70 ± 3.5 | 42.90 ± 3.2 | 49.43 ± 3.6 | 43.06 ± 1.7 | 39.86 ± 2.4 | | ShuffleFaceNet 1 *。*5× + LDA | 59.79 ± 5.3 | 54.21 ± 3.3 | 51.44 ± 3.8 | 64.33 ± 3.5 | 59.64 ± 2.0 | 57.57 ± 3.0 |   表6.在视频到视频和视频到图像场景中，针对REP-YTF开放集识别协议的YouTube Face数据库的性能比较。结果报告为等级1和FAR = 1％时的平均DIR（％）。   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 方法 | 视频到视频 | |  |  | 视频到图像 | | | 运算（0.2） | 运算（0.5） | 运算（0.9） | 运算（0.2） | 运算（0.5） | 运算（0.9） | | LBinVF 2 + LDA | 37.70 ± 3.5 | 30.93 ± 1.6 | 29.85 ± 1.3 | 31.55 ± 3.8 | 24.46 ± 1.7 | 24.01 ± 1.4 | | VGG-Face + LDA | 60.60 ± 3.2 | 54.59 ± 1.6 | 52.33 ± 1.3 | 53.87 ± 3.3 | 47.49 ± 1.8 | 45.02 ± 1.1 | | Dlib + LDA | 71.91 ± 2.0 | 66.53 ± 1.7 | 64.18 ± 0.8 | 56.78 ± 3.4 | 50.90 ± 2.7 | 48.33 ± 2.1 | | ShuffleNetV2 \* 1 *。*5× + LDA | 82.65 ± 2.4 | 79.22 ± 1.4 | 78.03 ± 1.2 | 76.88 ± 2.4 | 72.38 ± 1.4 | 71.20 ± 1.2 | | ShuffleFaceNet 1 *。*5× + LDA | 86.83 ± 2.1 | 85.52 ± 1.1 | 84.61 ± 1.0 | 84.40 ± 2.1 | 81.89 ± 1.1 | 80.36 ± 1.0 |   表7.视频到视频和视频的REP-YTF封闭集识别协议在YouTube Face数据库上的第1级识别率 |

视频到图像场景。

### 轻量级人脸识别挑战

我们对最近发布的ICCV 2019轻型人脸识别（LFR）挑战[ 15 ] 进行了另外的实验。具体来说，我们参加了Track 1，它要求float32解决方案，其计算复杂度小于1G FLOP，模型大小最大为20 MB，特征尺寸最大为512。选定的ShuffleFaceNet 1 *。*基于第4.1节中的验证结果的5x 模型可以满足此轨道的要求。评估该模型并将其与两个测试集中提供的基准MobileFaceNet进行比较：大型图像和大型视频。

Trillion-Pairs数据集（deepglint-light）用作大规模图像测试集，其中包含来自LFW名称列表中名人的5.7K身份的约274K面部图像和来自Flickr的1.58M面部图像作为干扰因素。在大型视频测试仪的情况下，将使用iQIYI-VID测​​试仪（iQIYI-light），其中包括20K个具有10K身份的视频。将所有测试图像预处理为与训练数据相似的112 × 112 的大小。

表8 显示了在给定的错误接受率下，在LFR挑战中获得的真实正率方面的结果。与MobileFaceNet基线相比，可以看到我们的ShuffleFaceNet架构在图像测试集上的改进，而性能与视频数据集上的基线模型所获得的性能非常接近。

## 速度和存储空间性能

在本节中，我们旨在评估所提议架构的速度和存储空间，以演示-

|  |  |  |
| --- | --- | --- |
| 方法 | 深光 | 爱奇艺灯 |
| MobileFaceNet基准 | 64.69 | 47.19 |
| ShuffleFaceNet 1 *。*5× | 75.31 | 44.55 |

表8.分别针对大型图像（deepglint-light）和视频（iQIYI-light）测试集以TPR @ FAR = 10 -8和TPR @ FAR = 10 -4报告的关于LFR挑战的结果。

确定其在实时应用程序或计算受限平台中的可行性。

我们首先将ShuffleFaceNets模型与ShuffleNetV2 1 的实际速度进行比较*。*5× 和LFR Challenge上提供的MobileFaceNet模型[ 15 ]。通过对Wild数据集中的Labeled Faces进行12,000张图像推断得出效率测量值。我们在四个不同的设备（包括CPU Intel i7-7700HQ（移动处理器），Nvidia Quadro P2000，Nvidia GeForce GTX 1050Ti Mobile和Nvidia GeForce GTX 1660Ti）上测量了这些模型的每个图像的推断时间。所得结果示于表9 。对于我们的ShuffleFaceNet架构，与准确性行为类似，随着复杂度的增加，推理时间也会增加。以ShuffleFaceNet 1 为参考*。*5× ，它被选择为具有更好的速度精度权衡的一种，我们可以看到它明显比ShuffleNetV2 1 快*。*5× 和MobileFaceNet模型，尤其是在CPU是MobileFaceNet最慢。此外，在不同的GPU上，与GPU相比，所有评估模型的推理时间均大大减少。但是，ShuffleFaceNet 1 *。*5倍于所有同类产品。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 方法 | 速度（毫秒） | | |  |
| 移动 | Quadro | 1050 | 1660 |
|  | 英特尔i7 | P200 | 钛 | 钛 |
| MobileFaceNet [ 15 ] | 62.4 | 5.5 | 7.3 | 3.3 |
| ShuffleNetV2 \* 1 *。*5× | 33.0 | 5.3 | 12.3 | 2.8 |
| ShuffleFaceNet 0 *。*5× | 12.0 | 1.4 | 2.0 | 0.7 |
| ShuffleFaceNet 1× | 23.8 | 3.0 | 3.3 | 1.2 |
| ShuffleFaceNet 1 *。*5× | 29.1 | 4.7 | 4.7 | 1.9 |
| ShuffleFaceNet 2× | 37.5 | 6.4 | 6.9 | 2.4 |

表9.不同设备中的推理时间比较。

表10 给出了以FLOP，参数数量和模型大小（以ShuffleFaceNet 1 为参考）的计算复杂性*。*与之前部分中测试过的几个最先进的脸部深度模型相比，该分辨率为5倍。可以看出，该提议是评估模型中最轻的提议之一。这以及上面讨论的快速推理速度，使我们的ShuffleFaceNet 1 成为可能*。*5倍高度适用于实时或计算受限的人脸识别应用。

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | 复杂 | #Param。 | 模型 |
|  | （FLOP） | （男） | 尺寸 |
| 脸网[ 28 ] | 1.6B | 7.5 | 30 |
| VGG脸[ 26 ] | 15G | 138 | 526 |
| 轻型CNN-4 [ 37 ] | 1.5克 | 4.1 | 26 |
| 轻型CNN-9 [ 37 ] | 1.0克 | 5.6 | 32 |
| 轻型CNN-29 [ 37 ] | 3.9克 | 12.6 | 125 |
| MobileFaceNet [ 2 ] | 4.398亿 | 1.0 | 4.0 |
| MobileFaceNet [ 15 ] | 933.3百万 | 2.0 | 8.2 |
| ShuffleFaceNet 1 *。*5× | 577.5百万 | 2.6 | 10.5 |

表10. Shuffle-的存储空间和复杂性比较

FaceNet 1 *。*使用某些最新的人脸识别模型可达到5倍。

# 结论

在本文中，我们开发了一种名为ShuffleFaceNet的轻量级卷积神经体系结构，以学习用于面部识别的强大功能。在考虑的四个复杂性级别中，ShuffleFaceNet 1 *。*5x 型号表现出最佳的速度精度折衷。在不同的人脸识别基准上进行的广泛实验表明，该提案取得了最新技术成果，保持了极高的效率。特别地，考虑到相关架构的复杂性，通用CPU的大小和速度始终都更好。ShuffleFaceNet 1 *。*5×具有约10MB的模型大小和29 ms的推理CPU时间，这支持其在实时和移动人脸识别应用程序中的实用价值。

1. 原文

提供更好的翻译建议